



Australian Bureau of Statistics

1351.0.55.163 - Research Paper: Synthetic Microdata - A Possible Dissemination Tool, Oct 2018

Latest ISSUE Released at 11:30 AM (CANBERRA TIME) 15/10/2018 First Issue

Summary

Executive Summary

EXECUTIVE SUMMARY

Enhancing microdata access is one of the strategic priorities for the Australian Bureau of Statistics (ABS) in its transformation program. However, balancing the trade-off between enhancing data access and protecting confidentiality is a delicate act. The ABS could use synthetic data to make its business microdata more accessible for researchers to inform decision making while maintaining confidentiality. Australian businesses in some industries are characterised by oligopoly or duopoly. This means the existing microdata protection techniques such as information reduction or perturbation may not be as effective as for household microdata.

The research focuses on addressing the following questions: Can a synthetic data approach enhance microdata access for the longitudinal business data? What is the utility and protection trade-off using the synthetic data approach? The study compares confidentialised input and output approaches for protecting confidentiality and analysing Australian microdata from business survey or administrative data sources.

The preliminary results show that synthetic data can be a possible dissemination tool to make more business microdata accessible while ensuring confidentiality. The analysis shows that the confidentialised input approach provides more protection than the confidentialised output approach in this particular setting - one percent sample file of business microdata. This is partly because the researchers have access to the microdata so there is a stronger need to add more noise for protection. The amounts of utility loss from synthetic data and perturbation approaches are comparable because the estimated coefficients are similar. Synthetic data could be a possible approach for the ABS to consider to enhance access to business microdata.

About this Release

This research explores if synthetic data is a possible approach to make ABS business microdata more accessible for researchers while maintaining confidentiality. The analysis compares confidentialised input and output approaches for protecting confidentiality and analysing Australian microdata from business survey or administrative data sources. It addresses the following research questions: Can a synthetic data approach enhance microdata access for the longitudinal business data? What is the utility and protection trade-off using the synthetic data approach?

History of Changes

23/08/2019 Improved the chart formats and description of methods.

© Commonwealth of Australia

All data and other material produced by the Australian Bureau of Statistics (ABS) constitutes Commonwealth copyright administered by the ABS. The ABS reserves the right to set out the terms and conditions for the use of such material. Unless otherwise noted, all material on this website – except the ABS logo, the Commonwealth Coat of Arms, and any material protected by a trade mark – is licensed under a Creative Commons Attribution 2.5 Australia licence